Data Analytics

Logistic Regression

Name : Modem Praveen   Roll no : 17BCS035

Dataset: Telecom data

Objective: Finding Churn Column (yes/no)

Read File: Churn_data <- read.csv(file.choose())

Structure : str(Churn_data)

Output:

```
'data.frame':	3333 obs. of  11 variables:
 $ Churn          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ AccountWeeks   : int  128 107 137 84 75 118 121 147 117 141 ...
 $ ContractRenewal: int  1 1 1 0 0 0 1 0 1 0 ...
 $ DataPlan       : int  1 1 0 0 0 0 1 0 0 1 ...
 $ DataUsage      : num  2.7 3.7 0 0 0 0 2.03 0 0.19 3.02 ...
 $ CustServCalls  : int  1 1 0 2 3 0 3 0 1 0 ...
 $ DayMins        : num  265 162 243 299 167 ...
 $ DayCalls       : int  110 123 114 71 113 98 88 79 97 84 ...
 $ MonthlyCharge  : num  89 82 52 57 41 57 87.3 36 63.9 93.2 ...
 $ OverageFee     : num  9.87 9.78 6.06 3.1 7.42 ...
 $ RoamMins       : num  10 13.7 12.2 6.6 10.1 6.3 7.5 7.1 8.7 11.2 ...
```

Splitting data as train and test data

split <- sample.split(Churn_data, SplitRatio = 0.7)

train <- subset(Churn_data, split== "TRUE")

test <- subset(Churn_data, split== "FALSE")

Training test data with Logistic regression model

```
logit_model <- glm(Churn ~ ., data = train, family = "binomial")
summary(logit_model)
```

…………………………………………………………………………..

Output:

```
Call:
glm(formula = Churn ~ ., family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7909  -0.5030  -0.3438  -0.2066   3.0779

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -6.263187   0.711442  -8.804  < 2e-16 ***
AccountWeeks     0.001619   0.001778   0.910 0.362579
ContractRenewal -1.936251   0.184798 -10.478  < 2e-16 ***
DataPlan        -1.031622   0.715032  -1.443 0.149087
DataUsage        1.717443   2.470877   0.695 0.487009
CustServCalls    0.467424   0.049066   9.526  < 2e-16 ***
DayMins          0.042571   0.041724   1.020 0.307590
DayCalls         0.006207   0.003568   1.740 0.081927 .
MonthlyCharge   -0.175187   0.245154  -0.715 0.474855
OverageFee       0.422263   0.418142   1.010 0.312564
RoamMins         0.096792   0.027867   3.473 0.000514 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1692.5  on 2120  degrees of freedom
Residual deviance: 1349.6  on 2110  degrees of freedom
AIC: 1371.6

Number of Fisher Scoring iterations: 6
```

…………………………………………………………………………

Output is depending on Intercept , ContractRenewal , CustServCalls and RoamMins.So Iam training only those columns to my model.

```
logit_model <- glm(Churn ~ ContractRenewal + CustServCalls + RoamMins , data = train, family = binomial)
```

Predicting Churn with test data using before trained model

fitted.results <- predict(logit_model, test, type = "response")

some of the results:

```
  1          3          9         10         12         14         20
0.07701407 0.06188956 0.07002196 0.27491253 0.04910333 0.18476219 0.095646
13
        21         23         25         31         32         34
36
0.05494303 0.05060039 0.05372502 0.14893143 0.07871615 0.08775261 0.215047
88
        42         43         45         47         53         54
56
0.33155267 0.11156132 0.04754996 0.19571531 0.12300643 0.10619550 0.078145
00
        58         64         65         67         69         75
76
0.11940875 0.18835963 0.09822295 0.24601545 0.06553008 0.06097831 0.057768
67
        78         80         86         87         89         91
97
0.19787037 0.11021712 0.04729104 0.19537396 0.11880734 0.08342440 0.065663
09
        98        100        102        108        109        111         1
13
0.20421624 0.11000449 0.07385904 0.07589816 0.07757765 0.06799145 0.251339
75
       119        120        122        124        130        131         1
33
0.11156132 0.10177854 0.04254906 0.08649568 0.15372662 0.16867620 0.074252
10
       135        141        142        144        146        152         1
53
0.07210843 0.09162296 0.05453416 0.08712209 0.20976043 0.07757765 0.064093
29
```

Fixing threshold and making outputs as 0 and 1

fitted.results.new <- ifelse(fitted.results.new > 0.3,1,0)

table(test$Churn, fitted.results.new)

	0	1
0	958	57
1	137	60

These are my correct and wrong predictions

Accuracy

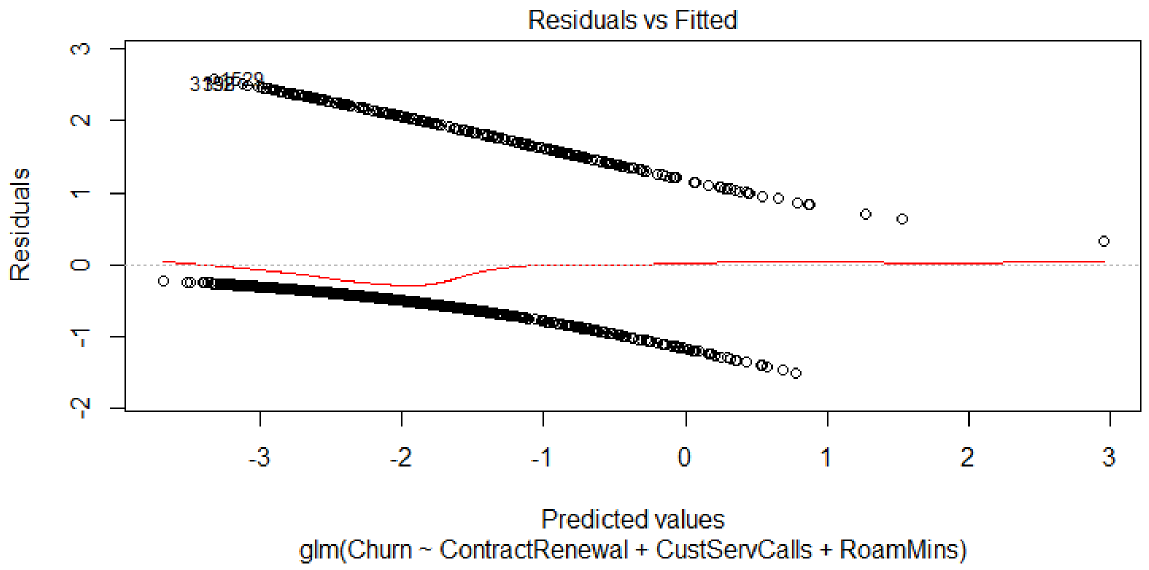
```
Error <- mean(fitted.results.new != test$Churn  )
print(paste('Accuracy =',1-Error))
```

Output: "Accuracy = 0.83993399339934"

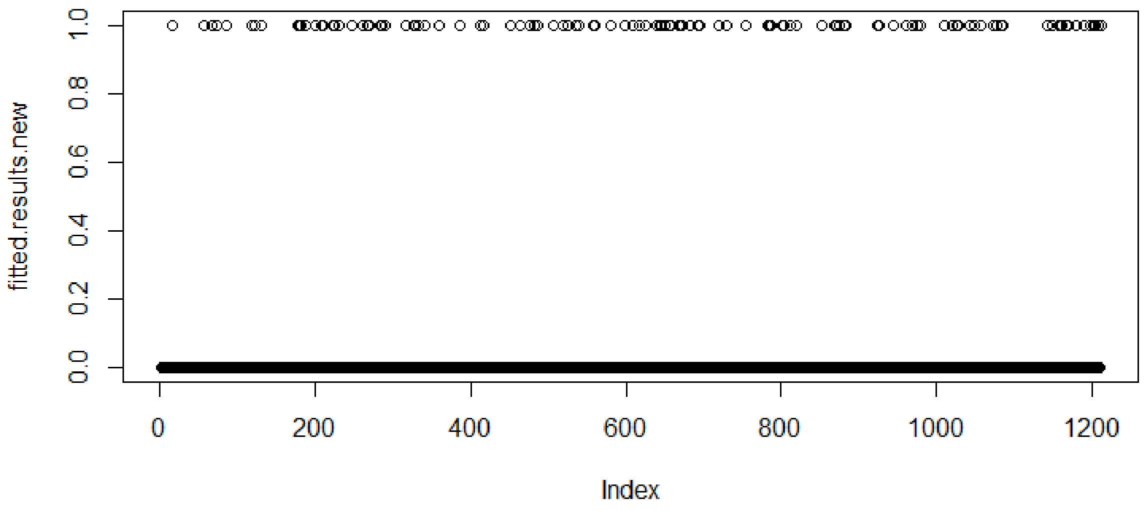
80% accuracy

Plots:

Model:



Binarised prediction values:


fitted.results.new
Index